**Assignment 3**

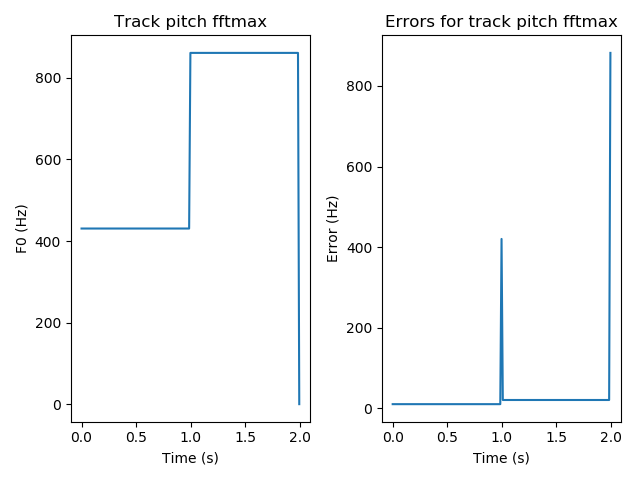
**Revisited: Fundamental Frequency Detection/Pitch Tracking**

Laney Light, Ning Yang, Zhuowen Lin

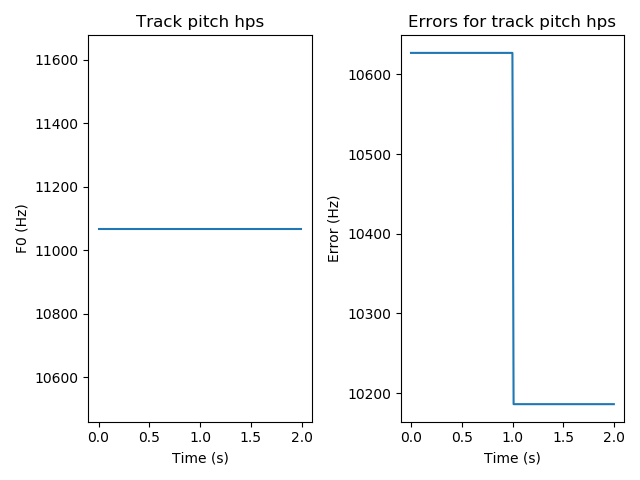
**A3.** If the block size=1024, the resolution is 44100 Hz/1024 = 43 Hz. Frequency reassignment, for example a constant Q transform, could be used to improve the resolution without changing the block size.

**E1.**

**Figure 1. Estimated F0 and Error for Max FFT Method (blockSize = 1024)**



**Figure 2. Estimated F0 and Error for HPS Method (blockSize = 1024)**

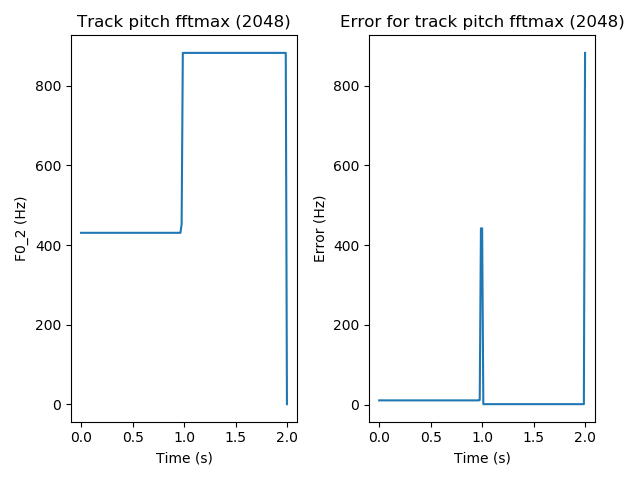


Discussion:

The HPS method fails for this signal because it is a pure sine wave. The HPS method uses pure multiplication weighting, so it only works for signals that have a full set of harmonics with sufficient magnitudes. For a pure sine wave like the input signal here with only one harmonic / one frequency, the multiplication of order 4 in HPS contains the multiplication of the only one harmonic and several zeros, giving result 0 and causing large difference from the ground truth fundamental frequency.

**E2.**

**Figure 3. Estimated F0 and Error for Max FFT Method (blockSize = 2048)**



Discussion:The peak in the error rate at the one second mark is slightly wider with a larger block size of 2048 than with a smaller block size of 1024.

**E3.**

**Table 1. Average performance metrics for Max FFT method**

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **RMS Error (Cents)** | **False Positive Rate** | **False Negative Rate** |
| Max FFT | 2801.85 | 0.99 | 0 |

**E4.**

**Table 2. Average Performance Metrics for HPS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **RMS Error (Cents)** | **False Positive Rate** | **False Negative Rate** |
| HPS | 7314.56 | 0.99 | 0 |

**E6.**

**Table 1. Average Performance Metrics by Method using Voicing Mask**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **Voicing Mask Threshold** | **RMS Error (Cents)** | **False Positive Rate** | **False Negative Rate** |
| ACF | -40 | 1081.86 | 0.17 | 0.01 |
| HPS | -40 | 6073.25 | 0.17 | 0.01 |
| Max FFT | -40 | 2176.67 | 0.17 | 0.01 |
| ACF | -20 | 3398.32 | 0.00 | 0.35 |
| HPS | -20 | 5767.53 | 0.00 | 0.35 |
| Max FFT | -20 | 3728.62 | 0.00 | 0.35 |

**Bonus:**

Based on the evaluation, ACF is the best choice overall for the implementation. So our approach is to based on the ACF with an voicing detection.